

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.
No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.
See full “License and Conditions of Use” notice for details.
In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.
Author: Carlos Daniel Borrego Romero

Annex D. Robust and Secure 2026-Compatible Protocol for Collaborative Studies on the Biological Layer of the Deterministic Quantum Hash Framework

Purpose

This annex defines a robust, secure, and scalable protocol for laboratories, hospitals, biobanks, universities, public agencies, and research organizations to perform new studies and analyses on the biological layer of the Deterministic Quantum Hash Framework in 2026, including use cases such as vaccines, mutations, molecular surveillance, translational research, and multi-site biological provenance studies.[file:202][web:274][web:275][web:281]

The protocol is designed to be compatible with the technological and regulatory landscape of 2026 by combining federated data analysis, controlled metadata harmonization, role-based governance, secure provenance capture, and AI-assisted integration without requiring unrestricted exchange of raw sensitive biological data.[web:269][web:272][web:277][web:283]

D.1 Scope and Design Goal

The biological layer of the framework is intended to generate a deterministic identifier for biological states and configurations using aggregated, structured, non-sensitive descriptors rather than raw genomic sequences or patient-level data.[file:202] This makes the framework particularly suitable for secure collaborative studies in which multiple institutions need to compare, track, validate, and interpret biological states while minimizing exposure of protected or regulated data.[file:202][web:274][web:276]

The core design goal is therefore to enable **collaborative biological analysis without unrestricted raw-data pooling**, while preserving provenance, reproducibility, auditability, and scientific interpretability.[web:272][web:277][web:280]

D.2 Security and Governance Principles

The protocol follows seven operational principles:[file:202][web:274][web:275][web:281]

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.

No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.

See full “License and Conditions of Use” notice for details.

In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.

Author: Carlos Daniel Borrego Romero

- **Data minimization:** raw human genomic or protected health information should not be moved unless specifically authorized and legally justified.[web:274][web:281]
- **Federated analysis first:** keep sensitive data at the source institution whenever possible and exchange only authorized intermediate or aggregated outputs.[web:272][web:277][web:283]
- **Role- and project-based access control:** access must be constrained by institutional role, study protocol, and approved project scope.[web:275]
- **Immutable auditability:** all access, transformations, and hash-generation events must be logged in secure non-editable audit trails.[web:275]
- **Schema-governed interoperability:** biological descriptors, metadata, and provenance fields must be standardized before cross-site computation.[file:202][web:278]
- **Human oversight:** AI may support harmonization and anomaly detection, but decisions affecting sensitive biological interpretation require expert review.[web:247][web:274]
- **Versioned reproducibility:** all operator definitions, entropy conventions, biological parameter sets, and serialization rules must be version controlled.[file:202]

D.3 Institutional Roles

The protocol requires explicit institutional roles:[web:275][web:281]

- **Data Custodian:** controls local biological data repositories and access authorization.
- **Scientific Principal Investigator:** defines study objectives and approves analytical interpretation.
- **Bioinformatics Lead:** manages pipelines for feature extraction, entropy estimation, and local model execution.
- **Security Officer / Compliance Officer:** verifies alignment with local security and legal requirements.[web:273][web:275]
- **Federation Coordinator:** manages cross-site schema alignment, study-wide identifiers, and approved inter-institutional exchange.
- **AI Integration Supervisor:** reviews AI-generated mappings, metadata harmonization outputs, and anomaly alerts.[web:247]

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.

No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.

See full “License and Conditions of Use” notice for details.

In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.

Author: Carlos Daniel Borrego Romero

- **Audit Reviewer:** periodically evaluates logs, access patterns, and reproducibility evidence.[web:275]

D.4 Biological Data Classes Covered by the Protocol

The protocol supports a layered classification of biological inputs:[file:202]

1. **Biospecimen descriptors:** sample type, tissue origin, collection protocol, storage history, quality-control indicators.[web:275][web:278]
2. **Molecular aggregate descriptors:** mutation counts, variant burden categories, lineage codes, expression-summary statistics, proteomic or metabolomic aggregate profiles.
3. **Experimental-condition descriptors:** vaccine formulation class, treatment arm, assay type, challenge condition, timepoint, dosage group.
4. **Population or cohort descriptors:** stratified non-identifying cohort statistics.
5. **Workflow provenance descriptors:** protocol version, software version, instrument version, operator version, and quality-control state.[file:202][web:275]

The protocol does **not** require direct circulation of raw genomes, raw patient records, or identifiable biospecimen data for most collaborative studies.[web:274][web:281]

D.5 Reference Study Types

The protocol is compatible with the following study categories:[file:202]

- **Vaccine response studies:** deterministic comparison of biological states across formulations, doses, timepoints, and laboratories.
- **Mutation and variant monitoring:** reproducible tracking of aggregated mutation or lineage states across sites.[web:269][web:272]
- **Biobanking research:** robust identity and provenance tracking across collection, storage, and distribution phases.[web:275][web:278]
- **Drug or intervention studies:** state comparison across conditions, response classes, and protocol variants.

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.

No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.

See full “License and Conditions of Use” notice for details.

In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.

Author: Carlos Daniel Borrego Romero

- **Translational research:** structured linkage between laboratory states, experimental conditions, and computational workflows.
- **Multi-institution surveillance:** federated comparative analysis of biological configurations without centralizing all sensitive data.[web:279][web:280]

D.6 Protocol Architecture

D.6.1 Local Secure Node

Each participating institution should operate a **Local Secure Node** that contains the protected biological data, local extraction pipelines, and the local Deterministic Quantum Hash processing components.[web:272][web:275][web:278] This node performs feature extraction, local normalization, entropy estimation, biological parameter encoding, metadata validation, and local generation of biological-layer descriptors.[file:202]

D.6.2 Federation Orchestrator

A **Federation Orchestrator** coordinates the collaborative study without requiring direct custody of all raw biological data.[web:269][web:277][web:283] Its responsibilities include study registration, schema distribution, protocol version control, approved query routing, aggregation policies, and cross-site comparison of permitted outputs.[web:274][web:277]

D.6.3 Provenance and Audit Service

A central or federated **Provenance and Audit Service** stores immutable records of study participation, node versions, schema versions, workflow execution events, and authorized exchanges.[web:275][web:278] This service should not store unrestricted raw biological data by default; instead it stores lineage, event metadata, and study-level reproducibility records.[web:275]

D.6.4 AI Harmonization and Monitoring Layer

An **AI-assisted Harmonization Layer** may be used to align metadata vocabularies, detect schema inconsistencies, flag improbable values, cluster similar study configurations, and generate explainability summaries for review.[web:243][web:247][web:249] This layer must operate under institutional governance, with reviewable outputs and constrained permissions.[web:247]

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.
No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.
See full “License and Conditions of Use” notice for details.
In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.
Author: Carlos Daniel Borrego Romero

D.7 Core Protocol Steps

D.7.1 Step 1. Study Registration

A collaborative study begins with registration of the scientific question, institutions, biological data classes involved, legal basis, approved study scope, and the exact descriptor families allowed for exchange.[web:274][web:276][web:281] This registration must assign a study identifier, protocol version, and risk category.[web:275]

D.7.2 Step 2. Schema and Vocabulary Alignment

Before analysis, the participating institutions must adopt a shared schema for biological descriptors, metadata fields, quality-control indicators, and causal or lineage codes.[file:202][web:278] Controlled vocabularies should be used for sample class, assay type, intervention class, mutation grouping, vaccine category, and protocol identifiers.[web:274][web:278]

D.7.3 Step 3. Local Data Classification and Risk Tagging

Each local institution classifies its data into categories such as raw protected data, restricted derived data, exchangeable aggregate descriptors, and public or semi-public study metadata.[web:275][web:281] Risk tagging determines which data may remain only local and which outputs are eligible for federation.[web:273][web:275]

D.7.4 Step 4. Local Feature Extraction

Each node computes authorized biological indicators from local data, such as variant burden summaries, lineage labels, expression heterogeneity scores, biospecimen quality indicators, experimental condition encodings, or cohort-level aggregate features.[file:202] These indicators are designed to represent the biological layer without exposing raw identifiable content.[file:202][web:274]

D.7.5 Step 5. Local Biological Descriptor Construction

The local node constructs the biological descriptor according to

$$H_q^{\text{bio}} = \mathcal{H}(E_{\text{bio}}, S_{\text{bio}}, T, \text{bio-parameters}, \mathcal{C}_{\text{bio}}),$$

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.
No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.
See full “License and Conditions of Use” notice for details.
In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.
Author: Carlos Daniel Borrego Romero

using locally approved effective scale descriptors, biological entropy estimators, temporal rules, causal or lineage constraints, and structured metadata.[file:202] At this stage, the continuous biological descriptor and the final deterministic hash are computed locally unless the study specifically authorizes a central computation on already approved derived data.[file:202]

D.7.6 Step 6. Federated Exchange of Approved Outputs

Only approved outputs are exchanged across sites, such as deterministic hashes, bounded descriptors, cohort aggregates, or encrypted intermediate statistics required for federated analysis.[web:272][web:279][web:283] Modern secure federated computation for genomics shows that multi-site analysis can be performed with strong confidentiality guarantees while keeping source datasets local.[web:272]

D.7.7 Step 7. Cross-Site Comparative Analysis

The federation layer compares approved outputs across institutions to identify concordant biological states, divergence patterns, temporal evolution, mutation clusters, vaccine-response profiles, or protocol-dependent differences.[file:202] This step may include secure aggregation, threshold queries, cluster analysis, or explainability reports generated from non-sensitive descriptors.[web:269][web:272]

D.7.8 Step 8. Review, Approval, and Publication Control

Before any cross-site result is disclosed, a scientific and compliance review verifies that outputs remain within the approved disclosure envelope and do not create re-identification or unauthorized inference risks.[web:274][web:281] Publication packages should include descriptor definitions, schema version, provenance summary, and permitted interpretation scope.[file:202]

D.8 Security Controls Fully Compatible with 2026

D.8.1 Access Control

The protocol requires role-based and project-based access control, consistent with biospecimen informatics best practices and security expectations for sensitive data systems.[web:275] Access must

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.
No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.
See full “License and Conditions of Use” notice for details.
In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.
Author: Carlos Daniel Borrego Romero

be granted according to least privilege, time-bounded approval, and protocol-specific authorization.[web:275][web:273]

D.8.2 Audit Logging

Every access to protected data, descriptor-generation event, model execution, export action, and AI-assisted transformation must be logged in a secure, permanent, non-editable audit trail.[web:275]

Logs should include actor identity, timestamp, dataset or node reference, operation type, and approval context.[web:275]

D.8.3 Federated Analysis and Secure Computation

For collaborative studies involving sensitive genomic or mutation-related analysis, the recommended default is federated analysis supported where needed by secure computation techniques such as homomorphic encryption or secure multiparty computation for approved aggregate operations.[web:272][web:283] This is fully compatible with the 2026 landscape and is particularly appropriate for multi-institution studies where legal and trust boundaries prevent unrestricted data pooling.[web:269][web:277]

D.8.4 Infrastructure Security Baseline

Institutions handling controlled genomic or biospecimen-related data should align systems with NIST-based security expectations and, where applicable, moderate baseline controls referenced in current genomic data-sharing practice.[web:273][web:275] Where human biospecimens or human genomic data are involved, policies that enhance security measures and controlled-access governance should be applied to the participating infrastructure.[web:270][web:281]

D.8.5 Data Segregation and De-Identification

Protected raw data, approved derived descriptors, and public study metadata must be stored in separate security zones.[web:275][web:281] Identifiable fields must be masked from incidental exposure, and descriptor exchange should use coded study identifiers rather than direct personal or sample identifiers.[web:275]

D.9 AI in the Protocol

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.

No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.

See full “License and Conditions of Use” notice for details.

In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.

Author: Carlos Daniel Borrego Romero

AI should be used as a **controlled analytical amplifier** rather than as an unconstrained autonomous actor.[web:243][web:247][web:249] In this protocol, AI may perform the following authorized functions:

- metadata harmonization across labs,
- anomaly detection in descriptors or workflows,
- ontology mapping,
- clustering of biological descriptor profiles,
- generation of explainability summaries for expert review,
- assistance in selecting candidate indicators for new study classes.[web:243][web:247][web:249]

AI should not automatically approve disclosure, override governance rules, or directly expose raw sensitive data to external model providers.[web:274][web:281] All AI-assisted outputs affecting scientific interpretation or exchange must remain reviewable and attributable.[web:247]

D.10 Scalability Strategy

The protocol is scalable because it is designed as a federation of repeatable local nodes rather than as a monolithic central repository.[web:269][web:277][web:279] Scalability is achieved through:

- schema versioning,
- reusable node software templates,
- shared policy registries,
- controlled descriptor families,
- asynchronous federated queries,
- institution-specific connectors under a common interoperability model.[file:202][web:274][web:277]

This makes the protocol suitable for small laboratory collaborations, national research networks, hospital consortia, vaccine-monitoring alliances, and global mutation-surveillance projects.[web:279][web:280]

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.

No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.

See full “License and Conditions of Use” notice for details.

In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.

Author: Carlos Daniel Borrego Romero

D.11 Output Artifacts of Each Study

Each collaborative study should generate the following artifacts:[file:202]

- study protocol identifier,
- schema and vocabulary version,
- list of participating nodes,
- descriptor definition set,
- local validation records,
- cross-site aggregation records,
- explainability report,
- audit summary,
- publication-control decision record.

These artifacts ensure that the biological layer remains scientifically interpretable and institutionally governable.[file:202][web:275]

D.12 Expected Scientific Value

A robust biological protocol for the Deterministic Quantum Hash Framework can enable reproducible tracking of biological states across vaccine studies, mutation monitoring, experimental pipelines, and translational research without forcing institutions into unsafe raw-data centralization.[file:202][web:272][web:279] Its strongest value lies in combining provenance, comparability, and secure collaboration in a way that is compatible with current federated genomics and biospecimen governance practices.[web:274][web:277][web:281]

D.13 Conclusions

A robust and secure protocol for the biological layer of the Deterministic Quantum Hash Framework is fully feasible in 2026 when built on federated analysis, controlled descriptor exchange, immutable audit logging, role-based governance, schema harmonization, and AI-assisted but human-supervised integration.[web:269][web:272][web:275][web:281] This design is scalable across laboratories and organizations because it keeps sensitive biological data local while enabling reproducible cross-site

This model and documentation are provided for non-commercial use only under the CC BY-NC-SA 4.0 license.

No commercial use, profiling, or exploitation of IP logs or trade secrets is permitted.

See full “License and Conditions of Use” notice for details.

In case of conflict or ambiguity, the most restrictive interpretation in favor of non-commercial use, user privacy, and protection of trade secrets shall prevail.

Author: Carlos Daniel Borrego Romero

analysis through approved deterministic descriptors and governed aggregate outputs.[file:202][web:277][web:283]

The protocol is especially well suited for collaborative studies involving vaccines, mutations, biospecimen research, and translational analysis because it strengthens provenance and scientific comparability without sacrificing security or regulatory compatibility.[web:274][web:279][web:280] In operational terms, it provides a realistic pathway for turning the biological layer of the framework into a deployable research infrastructure rather than a purely theoretical extension.[file:202]

LICENSE AND CONDITIONS OF USE.

The content is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license:

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

By using this content and model, you agree to:

Provide appropriate credit, include a link to the above license, and indicate if changes were made.

Not use the material or any derivative works for commercial purposes.

Distribute any adaptations or derivative works only under the same CC BY-NC-SA 4.0 license.

Not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Data logging and trade secrets policy.

The design, implementation, and deployment of this model, as well as any derivatives, do not authorize the collection, storage, or exploitation of IP addresses, unique device identifiers, or other technical traceability data for commercial profiling, targeted advertising, or any other economic exploitation.

Any system implementing this model must limit the logging and retention of IP addresses and other identifiable data strictly to what is necessary to:

(a) maintain technical security and service integrity; and

(b) comply with applicable legal obligations (for example, any minimum log retention required by law).

It is prohibited to use this model or its derivatives to:

capture, store, or exploit trade secrets, proprietary know-how, or confidential information of users or third parties for any purpose other than providing the technical service described in this documentation; or train or improve closed, commercial systems using data that contains confidential information of third parties without their prior, explicit, written consent.

---+EOD+---